Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 147 (2014) 213 – 222

Procedia
Social and Behavioral Sciences

ICININFO

The Data Exchange Model for Archiving (MEDONA)

Claire Sibille-de Grimoüard*

Service interministériel des Archives de France, 56 rue des Francs-Bourgeois, 75003 Paris, France

Abstract

In 2006, a Data Exchange Model for Archiving (MEDONA) was developed by the French Directorate of Archives (Ministry of Culture and Communication) and the Directorate-General for the modernization of the State (Ministry for Economic Affairs, Finance and Industry). The Model is currently being standardized by a working group of the French Association of Standardization including archivists of public archival agencies, storage services providers, software editors and IT consultants. MEDONA provides a model for the various specific transactions, which occur between the archival agencies and its partners (originating agencies, access requesters, other archival agencies, control authorities), i.e.: transfer; delivery; modification notification; destruction notification; restitution. Each transaction is modeled as a scenario for exchange of messages. These messages are formally defined as XML schemas.

Selection and peer-review under responsibility of the 3rd International Conference on Integrated Information.

Keywords: archival agencies; data exchange model; interoperability; modeling; OAIS; originating agencies; software editors; standardisation; UML; XML

1. Introduction

Data processing and office automation, already present in public services, in enterprises and in households for several decades, continue to develop. However, far from the aim of "zero paper" that the emergence of micro-computers had promised us, now we look into the objectives of efficiency, productivity improvement or services of better efficiency and better quality that data processing could bring.

The administration and the public services are taking an active part in this evolution of our society. A typical example is the take-off of electronic administration, which is constantly accelerating nowadays with different governmental programs (DISIC, 2012). No longer at stake now is the massive production of digital information

* Corresponding author. Tel.: + 33-1-40-27-60-15; fax: +33-1-40-27-66-42.
E-mail address: claire.sibille@culture.gouv.fr

Selection and peer-review under responsibility of the 3rd International Conference on Integrated Information.
doi:10.1016/j.sbspro.2014.07.154

nor the extreme automation of the various processes in order to reach the "zero paper". It's rather the faculty of systems to organise information and make it available rapidly in an easy to understand form. Electronic exchange plays an important role within a context where mainly the organisation efficiency and the exchange facilitation are looked for. The Data Exchange Model for Archiving or MEDONA is placed at this strategic level of the information system (Banat-Berger *et al.*, 2009; more information, is available through the website of the Service interministériel des Archives de France: http://www.archivesdefrance.culture.gouv.fr/seda/). It aims at facilitating the interoperability between the information system of an archival agency and the information systems of its partners: Originating Agency, Access Requester, Control Authority, and Transferring Agency. It provides a framework for the various specific transactions, which may occur: transfer, delivery, destruction, etc.

The Data Exchange Model for Archiving was developed in 2006 by the French Directorate of Archives (Ministry of Culture and Communication) and the General Directorate of the modernization of the State (Ministry of Budget) within the framework of a specific governmental program, dedicated to archival archiving (Banat-Berger, de Solan, 2008). It is currently being standardized by a working group of the French Association of Standardization, including archivists of public archival institutions, storage services providers, software editors and IT consultants. Currently, a draft of standard is open for comments (http://www.enquetes-publiques.afnor.org/information-et-communication-numerique/pr-nf-z44-022.html). The future national standard should be published by the end of 2013 (AFNOR, 2013; see also CINES, 2012).

2. Scope and key principles

2.1. Purpose and principles of development

The MEDONA standard provides a standardized framework for the various information exchanges (data and metadata) between the Archival agency and its partners. Interchanges between several archival agencies (integrated records management offices, public archival agencies, archiving providers) are also to be considered. The MEDONA standard defines activity diagrams and data models according to the UML standard, as well as in message definitions according to XML schemas.

The model presents useful elements for building applications above and below the messages. In particular, it defines the necessary data contained in these systems. This data will then be used to generate messages, by mapping between the model of the relevant databases and the XML schemas of the transactions envisaged in the MEDONA model. The model is generic and adaptable to all types of documents and data, both electronic and in paper format. It is compatible with any metadata standard: it provides a minimum technical metadata set but is open to any other any packaging standard.

Whereas other exchange standards, such as PAIMAS (ISO 20652: Space Data and Information Transfer Systems - Producer-Archive Interface - Methodology Abstract Standard), only cover the first stages of the ingest process defined by OAIS (Open Archival Information System), the situations that are covered by the MEDONA standard include: transfer, delivery, destruction, modification and restitution of digital documents or data between the transferring agency, archival agencies and third party entities. The MEDONA standard defines scenarios used by archival agencies and their partners in order to exchange information within the framework of these transactions. It also provides an implementation model using XML formalism for the exchanged messages. The aim of this work is to promote the interoperability between the information systems of public services and to allow for a better sharing in software developments (Jacobson, 2010).

2.2. Targeted audience

The standard is intended in particular for:

- the originating agencies of public records, such as ministries, decentralized State services, territorial authorities, public institutions;
- public archival agencies, in order to standardize electronic archive reception and delivery and to support the multi-site consultation portals;
- software developers who wish to comply to a standardized work frame when developing archiving modules;
- software developers of systems that manage and describe paper archives;
- consulting companies of exchange services working for public originating agencies and who may be required by these originating agencies to transfer documents to public archival agencies;
- storage service providers.

Actors with a managerial role will be able to rely on this model to:

- align the description of their archival process,
- set up the organization necessary for the management of information flows,
- refine the management of the contents and the checking of information items to add, suppress or deliver.

The model is already used: for the transfer of town decisions to local archives; for the transfer of files of Departmental Homes for Disabled Persons to local archives; for the transfer of public procurement files to other local archives (Moufflet, 2012).

2.3. Limits of the model

The data modeling is designed to be suitable for the handling and the description of both paper and electronic archives. However, the standard is limited to information exchanged between the various actors and is not concerned with the internal organisation of the actor's information systems. Information received in conformity with the exchange standard is destined to be handled by various components. But the components are not the object of this standard. The standard does not specify:

- how messages are to be transferred between the actors. In some contexts, protocols such as SFTP, FTP/S, HTTP/S will be relevant. In other contexts (e.g. for big data), a storage media exchange may be considered. Lastly, some businesses areas require secured or proprietary protocols. In particular, all messages exchanged over the network by the archival agency and its external partners can be secured by using appropriate protocols to ensure that the data recipient received all the data exchanged, that the actors are well known and that the data remain encrypted during the transfer over the network;
- how digital objects are transferred with the message (packaging format);
- a precise list of file formats for transfer of electronic records from a transferring agency to an archival agency;
- the naming conventions that may apply to the messages themselves;
- the information description model;
- the processing application. In some cases, processing along the way (transaction) will be relevant. In other cases (e.g. for big data, frequent exchanges or peak activity), batch will probably be preferred;
- how to remove information (dereferencing, data erasure or physical destruction of the data storage medium);
- data recovery;
- exchanges between a control authority and another;
- search and retrieval features.

All these aspects must be negotiated by the actors of the exchange and shall be determined by agreement, note or service contract between the actors when initiating the process.

2.4. Standardization references

From the very beginning it was decided to apply standards that would bring to our work a longevity, which would allow it to be shared with foreign partners (RGI, 2013). To achieve this, we relied on the following references:

- ISO 14721:2003 -- *Space data and information transfer systems – Open archival information system – Reference model*, better known as the OAIS (Open Archival Information System): this conceptual framework, established by the most important worldwide space centres, including the CNES (National Centre for Space Studies; in French: Centre national d'études spatiales), specifies the information objects, the metadata necessary for their preservation and management, their preservation and their delivery. The MEDONA standard has taken into account the definition of the actors, the exchanges and the information objects exchanged as defined in OAIS model;
- *ISO/IEC 19501:2005 -- UML* (Unified Modeling Language, version 1.4.2): a standard visual modeling language intended to be used for modeling business and similar processes, analysis, design, and implementation of software-based systems; the UML formalism has been chosen to express the MEDONA model.
- *ISO 20652:2006 -- Space data and information transfer systems - Producer-Archive interface - Methodology abstract standard* (PAIMAS): In MEDONA, the interactions which take place between an information producer and a deposit archive regarding the first stages of the ingest process were inspired by PAIMAS.
- *ISO/IEC 15489-1:2001 -- Information and Documentation - Records Management - Part 1: General*;
- *ISO 23081-1:2006 -- Information and documentation - Records management processes - Metadata for records - Part 1: Principles*;
- *ISO 23081-2:2007 -- Information and documentation - Records management processes - Metadata for records - Part 2: Conceptual and implementation issues*;
- Other standards: XML Schema and OWL

3. Description of the model

3.1. The actors

An actor represents a coherent set of roles. The actor should not be confused with the user (or individual) that can play different roles depending of the used transaction. For example, an agent that transfers information generated by his agency will serve as a Transferring agency during the transfer, as an Originating agency during the restitution and as an Access requester during his access requests.

The Archival agency is the actor receiving the transfer and is responsible for the management of the Archives transferred by the Transferring agencies and their delivery back to the Originating Agency and, depending on access issues, to Access Requesters. The Transferring agency refers to the actor that transfers an Archive to an Archival agency. The Originating agency refers to the actor that produced the archives, i.e. that created them or received them within the context of its activities. This may, or may not, be the same as the Transferring agency. Some messages are issued only after validation by a Control authority. In the French public sector, this control authority keeps the archives under strict scientific and technical control: Interdepartmental Service of French Archives, and directors of departmental archives at the local level. The term Access requester refers to any person or legal entity wishing to consult the Archives kept by the archival agency in respect of the legislation applicable in the field of archive delivery.

3.2. The types of exchanges

The main use cases considered (exchanges between actors) are:

- Archive Transfer: the archive transfer by a Transferring agency to an Archival agency in order to ensure an Archive's preservation. The transfer can be preceded by an Archive Transfer Request;
- Archive Delivery: delivery of an Archive by an Archival agency to an Access requester, with the authorization, if necessary, of the Originating agency and of the competent Control authority;
- Archive Modification Notification: information sent by an Archival agency to an Originating agency to inform it that Archives coming from it and preserved by the archival agency have been modified by the latter in order to ensure its correct preservation. The modifications could be concerned with the data themselves (for instance format conversion) or with the metadata (adding, correction, update).
- Archive Destruction: exchanges linked to the archive destruction, either at the request of the Archival agency with the authorization of the Originating agency and of the competent control authority, or at the request of the Originating agency, with the authorization of the competent control authority.
- Archive Restitution: Transfer of an Archive from an Archival agency back to the Originating agency.

3.3. The exchanged objects

The main objects exchanged during transactions are data objects, technical, descriptive, and administrative metadata. The types of these objects and the cardinality of their components are presented in class diagrams and specified in the XML schemas.

A package of data object (DataObjectPackageType) combines:

- a data content, i.e. the digital or physical object(s) that are the main concern of the long-term preservation;
- representation information, i.e. the information, which translates the data content into more explicit concepts (for instance, the definition of the ASCII code describes how a sequence of bits (a data content) is converted into characters);
- information for long-term preservation, i.e. information necessary to a good quality preservation of the information content (information on origin, identification, integrity, context).

The Model makes a distinction between:

- binary data objects (BinaryDataObjectType): for instance an electronic file, i.e. a named and ordered sequence of bytes that is known by an operating system;
- physical data objects (PhysicalDataObjectType): for instance a file, a box, a CD-Rom, etc.

The Management Metadata (ManagementMetadataType) include the following information:

- Archival Agreement: agreement (convention, contract) or regulation used as the framework for the relationships between the transferring agency and the archival agency;
- Archival Profile: rules for building the archive according to the type of documents or the type of application involved;
- Service Level: service level requested (e.g. availability, security...), referring to the various levels envisaged by the Archival Agreement;
- Access rules: information necessary for implementing procedures for the access to data objects and for expressing whether it is required to set up access restrictions;
- Appraisal rules: information needed to manage the data lifecycle to indicate a duration beyond which a data object is deleted or kept without time limit.

Descriptive metadata (DescriptiveMetadataType) include information for long-term preservation (data origin, description, date, keywords, etc.). These descriptive metadata, depending on the practices of the exchange partners, can follow different models.

4. Modeling

The formalism used to express the model is the formalism of UML (Unified Modeling Language). Three types of diagrams are used:

- The use case diagrams provide an overview of the system by representing the actors and their actions on the system;
- The sequence diagrams show every application case and provide a temporal representation of the progress of each action. These diagrams show the scenarios of actors;
- The class diagrams are used to provide a set of properties to all items handled when scenarios are implemented by actors.

4.1. Use case diagrams

Five main use cases occur between the archival agency and its partners: Transfer, Delivery, Archive Modification, Destruction and Restitution (see figure 1).

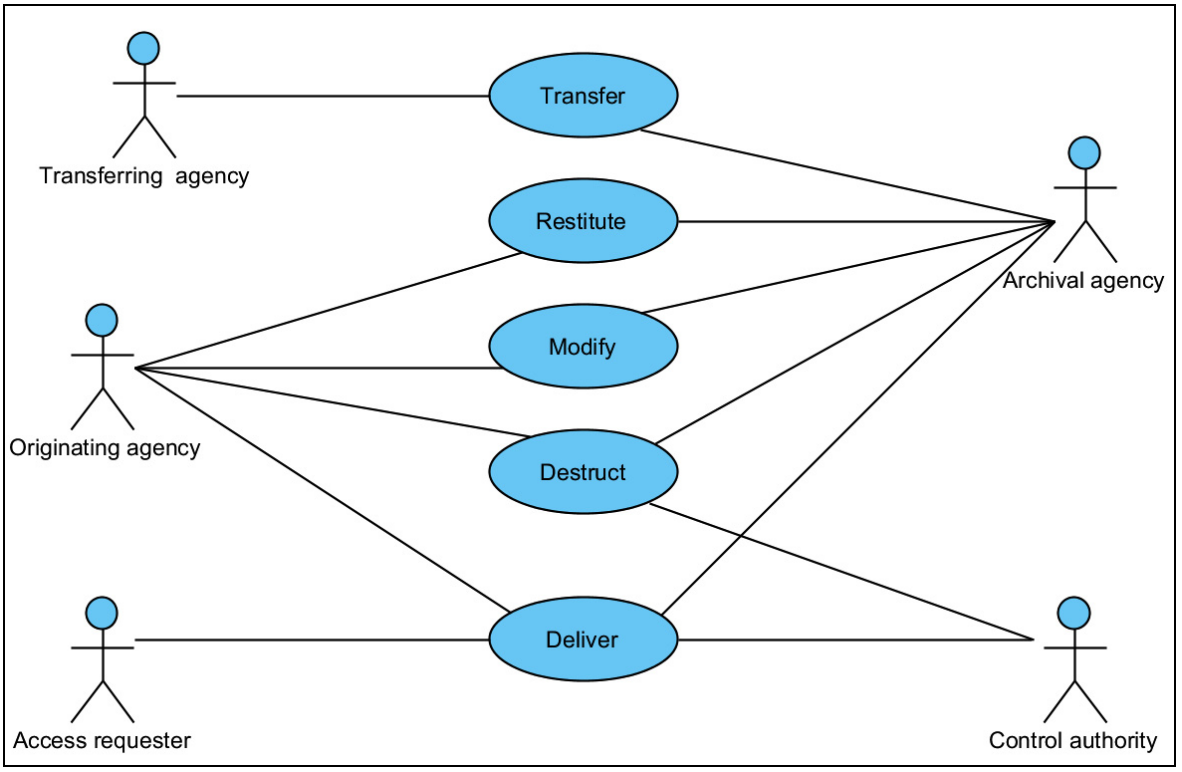

Fig. 1. The five main use cases

For instance (see figure 2), during transfer, the transferring agency transmits to the archival agency a set of information concerning the transfer itself (identification of the transferring agency and of the archival agency, the type of convention (or service contract) agreed between these two parties, level of description…) in addition to the

information on the objects to be preserved. If the objects are digital, the objects themselves can be attached to the transfer.

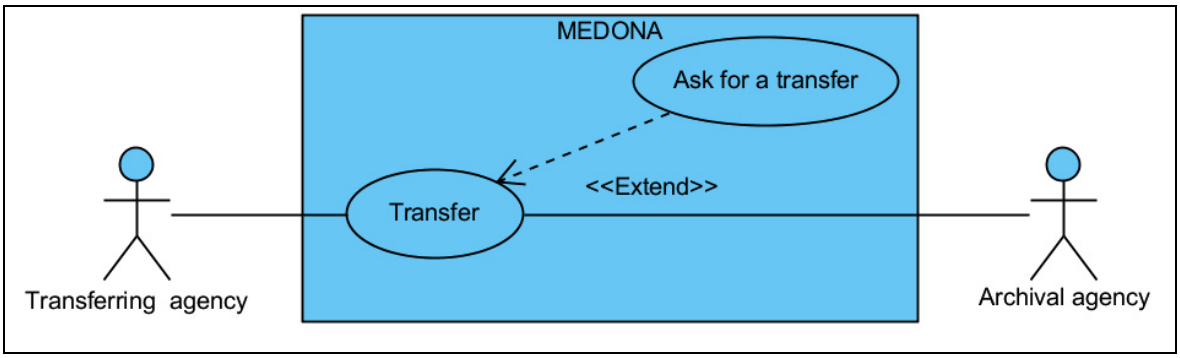

Fig. 2. Transfer use case

4.2. Sequence diagrams

The sequence diagrams describe the dialogue between actors in the context of a transaction. These diagrams identify messages that are sent by actors and describe the sequence of these messages. They cover only the exchange of messages and not the procedural aspects. These diagrams generally link together four messages: a request, an acknowledgment of receipt of the request, a response to the request and a receipt of the response.

For instance (see figure 3), transfer request allows the transferring agency to check with the archival agency that the planned transfer is acceptable by sending just the metadata for agreement. This request is optional. It is immediately acknowledged, then, in asynchronous context, by a notification from the archival agency that the transfer proposal is accepted or rejected.

The Transfer exchange transfers an archive from the Transferring agency to an Archival agency. The Transfer message comprises the archives to be transferred and their metadata. The transfer message is addressed to the archival agency by the transferring agency. When sent through a network, the originating agency receives Transfer acknowledgement immediately after the end of the transmission. The archival agency then checks that the transferred archives meet all the conditions specified in the convention or the service contract previously accepted by both parties. Either an archive acceptance notification or an error warning will be sent. The message for the acceptance includes the metadata of the transferred archives to confirm what the archival agency received. When an error warning is sent, the transferring agency is requested to send an acknowledgement of this warning. All the exchanges occurring during the archive transfer can be signed.

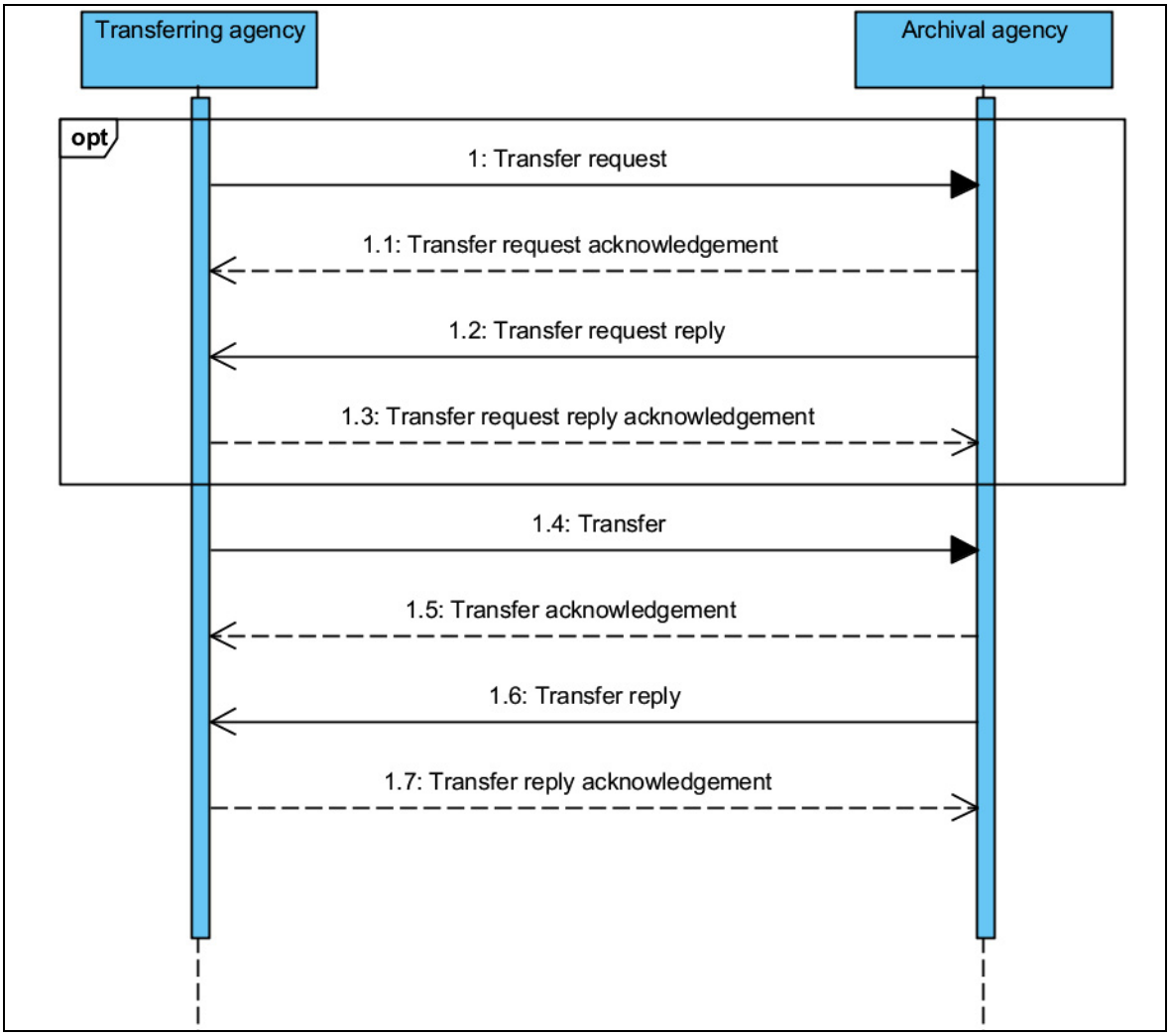

Fig. 3. Transfer exchange sequence

4.3. Class diagrams

Class diagrams describe the structure of messages that actors exchange in the framework of transactions, as well as the structure of the objects handled within these messages. For instance (figure 4), the Organization class enables us to describe the actors that interact in the exchanges (transferring agency, archival agency, originating agency...). They must be identified (Identifier). In addition, technical identifiers (xml:id) can be used to provide a reliable mechanism for managing references within the same XML document.

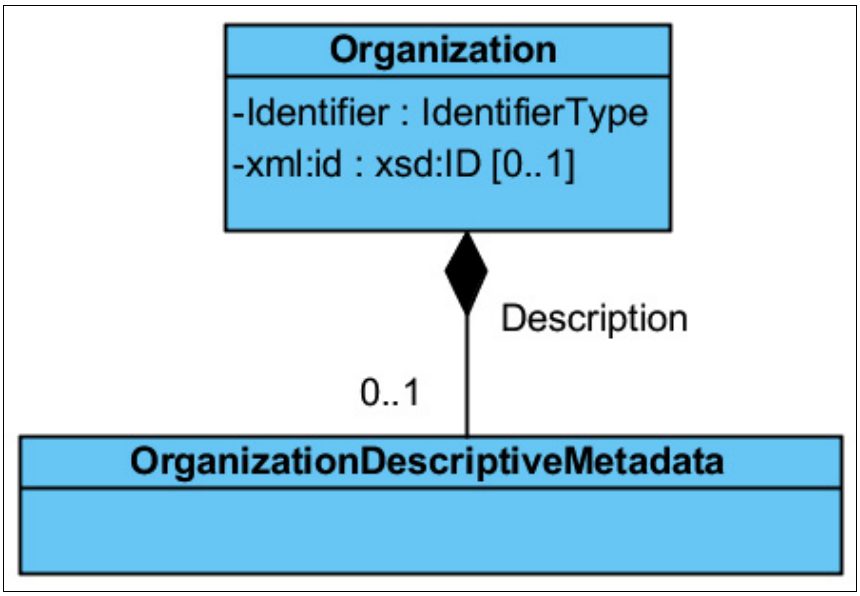

Fig. 4. Organization Class

5. Conclusion

At French level, the data exchange model for archiving has been incorporated into the Recommendation Framework for Interoperability established by the order n° 2005-1516 of 8 December 2005 relating to electronic exchanges between users and administrative authorities and between administrative authorities (DGME, 2009; see also DISIC, 2012). This framework sets the standards, which have to be used by the administrative authorities.

But the market of software is international, just like the one of data hosting and data transfer services. In addition, archives have similar functions throughout the world. An international standard would be therefore very highly useful. During the last meeting of the ISO/TC46/SC 4 (ISO sub-committee responsible for standards in the area of technical interoperability) at Paris, in June, we introduced the process of standardizing this protocol for the exchange of data between an archival institution and its partners. A resolution was adopted aiming to ask volunteers in SC 4 and SC 11 to review the AFNOR draft standard for consideration as a possible new work item.

Lastly, we plan to specify how public agencies should use this standard. In particular, it is advisable to specify the standards to be used regarding descriptive, administrative and technical metadata and how to use these standards (taking into account their own evolutions). It is also advisable to specify parts of the model (i.e. the model of intellectual archival description, the description model about actors and the description model about accessibility). For instance, regarding archival description, the model should refer to archival metadata element sets such as EAD, and there would be a need to standardize them too.

References

AFNOR (Association française de normalisation) (2013). Projet de norme MEDONA sur la modélisation des échanges de données pour l'archivage. Available through: http://www.afnor.org/liste-des-actualites/actualites/2013/juillet-2013/projet-de-norme-medona-sur-la-modelisation-des-echanges-de-donnees-pour-l-archivage-donnez-votre-avis. Accessed 29 July 2013.

Banat-Berger, F., de Solan, O. (2008). France's approach to data exchange. The data exchange standard for archiving. Available through: http://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/desolan_presentation_daf_pilae_20080422.pdf

Banat-Berger, F., Huc, C., Duplouy, L. (2009). *L'Archivage numérique à long terme, les débuts de la maturité?* Paris: La Documentation française

CINES (Centre informatique de l'enseignement supérieur) (2012), Publication of a new version V1.0 of SEDA (Data Exchange Standard for Archiving) national standard. http://www.cines.fr/spip.php?article1029. Accessed 29 July 2013.

DGME (Direction générale de modernisation de l'Etat) (2009). Référentiel général d'interopérabilité: http://references.modernisation.gouv.fr/standard-dechange-de-donnees-pour-larchivage-seda. Accessed 29 July 2013.

DISIC (Direction interministérielle des services d'information et de communication) (2012). Archivage électronique. Un nouveau domaine d'expertise au service de la gouvernance des systèmes d'information. Available through: http://references.modernisation.gouv.fr/sites/default/files/DISIC_AE_Guide_bonnes_pratiques_0.pdf. Accessed 29 July 2013.

ISO 14721–- Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. Geneva: International Organization for Standardization, 2012

ISO 15489 – Information and documentation – Records management, parts 1 and 2. Geneva: International Standards Organization, 2001.

ISO 20652 – Space data and information transfer systems - Producer-Archive interface - Methodology abstract standard (PAIMAS). Geneva: International Organization for Standardization, 2006

ISO 23081 – Information and documentation -- Records management processes -- Metadata for records, parts 1 and 2. Geneva: International Standards Organization, 2006-2007.

ISO/IEC 19501 – Information technology -- Open Distributed Processing -- Unified Modeling Language (UML) Version 1.4.2. Geneva: International Organization for Standardization, 2005

Jacobson, M. (2010). Le standard d'échange de données pour l'archivage (SEDA), version 0.2. Available through: http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20100121seda.pdf

Moufflet, J.-F. (2012). Utilisation du standard d'échange de données pour l'archivage. Available through: http://www.archivesdefrance.culture.gouv.fr/seda/formation/presentation_profil.pdf

SIAF (Service interministériel des archives de France). Standard d'échange de données pour l'archivage (SEDA). Available through: http://www.archivesdefrance.culture.gouv.fr/seda/. Accessed 29 July 2013.